



## ORIGINAL RESEARCH ARTICLE

# Evaluating Social Media Networks in Medicines Safety Surveillance: Two Case Studies

Preciosa M. Coloma<sup>1</sup> · Benedikt Becker<sup>1</sup> · Miriam C. J. M. Sturkenboom<sup>1</sup> · Erik M. van Mulligen<sup>1</sup> · Jan A. Kors<sup>1</sup>

Published online: 5 August 2015

© The Author(s) 2015. This article is published with open access at [Springerlink.com](http://Springerlink.com)

## Abstract

**Introduction** There is growing interest in whether social media can capture patient-generated information relevant for medicines safety surveillance that cannot be found in traditional sources.

**Objective** The aim of this study was to evaluate the potential contribution of mining social media networks for medicines safety surveillance using the following associations as case studies: (1) rosiglitazone and cardiovascular events (i.e. stroke and myocardial infarction); and (2) human papilloma virus (HPV) vaccine and infertility.

**Methods** We collected publicly accessible, English-language posts on Facebook, Google+, and Twitter until September 2014. Data were queried for co-occurrence of keywords related to the drug/vaccine and event of interest within a post. Messages were analysed with respect to geographical distribution, context, linking to other web content, and author's assertion regarding the supposed association.

**Results** A total of 2537 posts related to rosiglitazone/cardiovascular events and 2236 posts related to HPV vaccine/infertility were retrieved, with the majority of posts representing data from Twitter (98 and 85 %,

respectively) and originating from users in the US. Approximately 21 % of rosiglitazone-related posts and 84 % of HPV vaccine-related posts referenced other web pages, mostly news items, law firms' websites, or blogs. Assertion analysis predominantly showed affirmation of the association of rosiglitazone/cardiovascular events (72 %;  $n = 1821$ ) and of HPV vaccine/infertility (79 %;  $n = 1758$ ). Only ten posts described personal accounts of rosiglitazone/cardiovascular adverse event experiences, and nine posts described HPV vaccine problems related to infertility.

**Conclusions** Publicly available data from the considered social media networks were sparse and largely untrackable for the purpose of providing early clues of safety concerns regarding the prespecified case studies. Further research investigating other case studies and exploring other social media platforms are necessary to further characterise the usefulness of social media for safety surveillance.

## Key Points

The growing popularity of online communities and social media networks is stimulating exploration of these sources for pharmacovigilance purposes.

The potential value of mining data from social networks appears to be greater for measuring awareness regarding emerging safety issues.

Further research investigating other case studies (including prospective investigations) and exploring other social media platforms are necessary to further characterise the usefulness of social media for pharmacovigilance.

**Electronic supplementary material** The online version of this article (doi:[10.1007/s40264-015-0333-5](https://doi.org/10.1007/s40264-015-0333-5)) contains supplementary material, which is available to authorized users.

✉ Preciosa M. Coloma  
[p.coloma@erasmusmc.nl](mailto:p.coloma@erasmusmc.nl)

<sup>1</sup> Department of Medical Informatics, Erasmus University Medical Center, PO Box 2040, 3000 CA Rotterdam, The Netherlands

## 1 Introduction

The past decade has brought forth enormous growth and popularity of online communities and social networks, greatly expediting information exchange from one corner of the world to another. The concept of blogging has allowed virtually anybody with Internet access to post his or her views and experiences on any topic at any time. Whilst the value of such online conversations has been exploited mostly by commercial enterprises to promote product improvement and innovation, healthcare has not been immune to this phenomenon of public engagement [1–3]. In the same spirit of eliciting greater patient participation, several investigators have begun to explore what social media can offer in terms of medicines safety surveillance [4–6]. Reporting of individual cases of suspected adverse drug reactions (ADRs) to regulatory authorities, mostly by physicians or other healthcare professionals, remains the cornerstone of pharmacovigilance. However, spontaneous reporting systems are hampered by various limitations, the most important of which is under-reporting [7, 8].

Because social media represent secondary data, i.e. data that are not originally intended for surveillance, there are challenges to overcome with respect to terminology, traceability and reproducibility. Apart from these technical challenges, practical policy guidelines are lacking on how potential safety signals from social media should be handled in the current regulatory framework. Although the US FDA has released two guidance documents on the use of social media platforms for presenting benefit/risk information on prescription drugs and medical devices [9], these documents are more concerned with product promotion than surveillance and “do not establish legally enforceable rights or responsibilities” [10]. The European Medicines Agency (EMA) guideline on good pharmacovigilance practices (Module VI) [11] provides provisions on how to deal with information on suspected adverse reactions from the Internet or digital media, and hold marketing authorisation holders (MAHs) responsible for reviewing websites under their control for valid cases and reporting them accordingly, although there is no requirement to trawl Internet sites not under the control of the MAH. To date, there are no standard methodologies to mine user-generated data from social media for pharmacovigilance. In this study we sought to evaluate the potential contribution of mining social media networks for pharmacovigilance using examples of drug–event associations that have been flagged as potential signals: rosiglitazone and cardiovascular events (i.e. stroke and myocardial infarction), and human papilloma virus (HPV) vaccine and infertility.

## 2 Methods

### 2.1 Data Sources

Postings were collected from three of the most widely used social media networking platforms (Facebook, Google+, and Twitter) using their respective search application programming interfaces (APIs). The search APIs return a set of public messages from the social network that match the query keywords. For each message the content is provided, together with additional information about the message itself (date and content), about the status in a conversation (repost or reply), and about the author (account name and location). The results are encoded in machine-readable format (JavaScript Object Notation [JSON]) for integration into custom application. Messages were obtained from as far back as available until 25 September 2014. Only English-language posts were considered. Facebook provides only messages from the preceding month using their search API, while the search API of Google+ obtains messages dating back to its establishment in 2011. The search API of Twitter is restricted to a time window of approximately 1 week. In order to supplement the Twitter data obtained via its search API, an additional search engine, Topsy (<http://topsy.com/>) was used. Topsy is a real-time search engine for posts and shared content on social media, primarily on Twitter and Google+. Topsy has complete coverage of historical messages and has indexed every (public) tweet ever posted since 2006. As of this writing, Topsy was a Certified Reseller of Twitter’s data. For this particular study, only Twitter-related posts were retrieved via the free analytics service of Topsy.com. No Facebook or Google+ posts were retrieved in Topsy.

### 2.2 Case Studies

Usefulness of the above social media platforms for safety surveillance was evaluated using two examples of drug–adverse event associations that have previously been flagged as potential safety signals: (1) rosiglitazone and cardiovascular events (i.e. stroke and myocardial infarction); and (2) HPV vaccine and infertility. These two case studies were chosen because they represent associations that have triggered controversies and thus are likely to have been the subject of media attention as well as online discussions. Furthermore, the case studies involve different types of agents that are used by different subsets of the population under different circumstances, thus allowing investigation of diverse scenarios.

For each case study, data were queried for co-occurrence of the drug/vaccine of interest and the event of interest within the same post or tweet. Search queries were

constructed using all possible drug–event keyword combinations (the keywords used are provided in the Appendix, available as electronic supplementary material). Event-related keywords consisted of clinical terms from the Unified Medical Language System (UMLS), as well as known abbreviations and layman’s terms (see Appendix). Drug-related keywords consisted of international nonproprietary names and trade names.

### 2.3 Assessment of Suitability for Use in Safety Surveillance

Relevant posts were tallied and analysed with respect to geographical distribution, context, and linking to other web content. The country of origin of a message was automatically determined from the location information about the author. When the country was not available in a designated data field, it was manually identified from the available location information by means of a list of names of countries, regions and cities. The frequency of message propagation (i.e. reposts or retweets) was calculated. The content of all posts were reviewed one by one to determine whether there was reference to a person’s actual experience of having the (adverse) event of interest in relation to exposure to the drug (or vaccine) of interest. It was not the intention to assign or assess causality, but rather to describe the context of how the drug–event relationship is described. Posts were likewise analysed with respect to the author’s assertion of the purported association between the drug (or vaccine) of interest and the event of interest. Somewhat analogous to sentiment analysis, assertion was judged as one of the following: (1) ‘affirmative’, if the post alluded to an affirmation of the association; (2) ‘negating’, if the post alluded to a negation of the association; or (3) ‘neutral’, if the post alluded to neither affirmation nor negation of the association. Manual review and annotation of the assertions was undertaken by a physician/pharmacist (PMC). In addition, key dates during which important communication or regulatory actions occurred were marked and compared with the timeline of the posts.

## 3 Results

### 3.1 Rosiglitazone and Cardiovascular Events

As shown in Table 1, we retrieved a total of 2537 posts related to rosiglitazone and cardiovascular events (i.e. stroke and myocardial infarction), with the overwhelming majority of posts (98 %) representing data from Twitter. Only two posts were retrieved on Facebook, while 41 posts were retrieved on Google+. Approximately 10 % of all posts were reposts or retweets. The country of origin (based

on the holder of the social network account) could not be automatically identified in 59 % of the posts; of the posts that could be identified, two-thirds were accounted for by the US, while the remaining one-third was distributed among 50 other countries or territories all over the world. Overall, 21 % of posts ( $n = 536$ ) had links to other web pages (see Table 2). News items comprised more than one-third of the web pages referenced ( $n = 196$ ), followed by law firms’ websites or advertisements ( $n = 157$ ) and blogs ( $n = 138$ ). There were 24 posts referring to health information websites intended for health professionals, 15 posts linking to scientific journals, four posts referring to a patient community website, one post linking to a hospital’s patient education website, and another linking to a YouTube video.

Assertion analysis carried out on all posts predominantly demonstrated affirmation of the association between rosiglitazone and cardiovascular events (72 %;  $n = 1821$ ), with the remainder more or less split between negating (13 %) and neutral (15 %). Most neutral posts were asking for further information or were otherwise not directly related to the drug–adverse event association. There were posts by lawyers or reporters explicitly soliciting cases ( $n = 12$ ), but there were also posts ( $n = 122$ ) ridiculing lawyers’ television commercials that asked patients who ‘died while taking the drug’ to call a particular number. Figure 1 shows the trend of assertions over time in relation to events in the timeline of the association of interest. The highest peak of affirmative posts occurred in February 2010. In this particular month, the US Senate Finance Committee released a report based on a 2-year inquiry of rosiglitazone, expressing concern that the “FDA has overlooked or overridden safety concerns cited by its own officials” [12]. The EMA’s suspension of rosiglitazone’s marketing authorisation in the EU, and the FDA’s restriction of access to the drug, coincided with the second peak of affirmative posts in September 2010, while the simultaneous publication in high-impact journals of two studies demonstrating increased cardiovascular risk with the use of rosiglitazone [13, 14] coincided with the peak in June 2010. The peaks in negating assertions paralleled those of the affirmative, with the greatest peak in affirmations observed in June–July 2010 (and a smaller peak in November 2013), reflecting the active online debate that was happening regarding the issue. Figure 1 also shows that in June 2013, negating posts actually outnumbered the affirmative posts; the results of the FDA-mandated re-evaluation of the rosiglitazone (RECORD) trial [15] became available online in June 2013. The peak of neutral posts seen in July 2011 represented posts about news of rosiglitazone being potentially useful for neuropathic pain (although the pertinent study [16] had already been published online 3 months earlier).

**Table 1** Overview of posts about rosiglitazone and cardiovascular adverse events across social media networking platforms

Social media platform	No. of posts (%)	No. of reposts (%)	No. of posts with links to other sites (%)	Earliest/latest date of retrieved post	Origin of post (country, based on account holder) <sup>a</sup>
Facebook	2 (0.1)	0	2 (100)	July 2014/August 2014	US (1) Unknown (1)
Google+	41 (1.6)	6 (15)	41 (100)	June 2012/August 2014	Unknown (31) US (9) Egypt (1)
Twitter	2494 (98.3)	250 (10)	493 (20)	May 2007/September 2014	Unknown (1461) US (682) India (53) UK, Canada (50 each) Indonesia (31) Other countries (167)
Total	2537	256 (10)	536 (21)		

<sup>a</sup> Where applicable, only the top five countries are given

**Table 2** Description of web pages referenced by posts about rosiglitazone and cardiovascular events

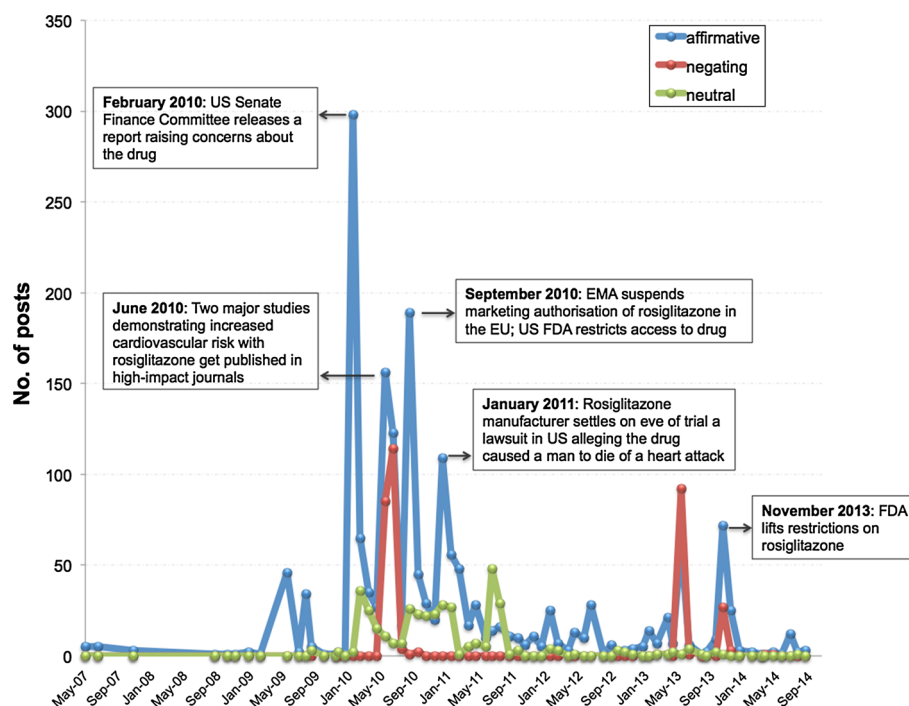
Category of linked web pages	Facebook [n = 2] (%)	Google+ [n = 41] (%)	Twitter [n = 493] (%)	Total [n = 536] (%)
News	–	8 (20)	188 (38)	196 (37)
Law firm's website or advertisement	1 (50)	17 (41)	139 (28)	157 (29)
Blog	–	13 (32)	125 (25)	138 (26)
Health reference for professionals	–	2 (5)	22 (5)	24 (5)
Patient community website	–	–	4 (1)	4 (<1)
Health education for patients	1 (50)	–	–	1 (<1)
Scientific journal	–	–	15 (3)	15 (3)
Video	–	1 (2)	–	1 (<1)

There were only ten posts that appeared to be about experiences of the drug–adverse event association of interest. Four posts involved the person posting the message himself or herself (one even claimed winning a legal case against the drug manufacturer); three involved somebody's brother-in-law, while there was one each for somebody's father, father-in-law, and grandmother. In addition, two posts referenced a patient community website that claimed 21,015 people reported to have had a heart attack while taking rosiglitazone (representing '32 % of all who reported side effects'). Interestingly, some posts ( $n = 20$ ) alleged other adverse events of rosiglitazone, such as leg pain, abdominal pain and eye pain (all of which are symptoms suggestive of end-organ complications of diabetes, the primary indication for the drug), while others ( $n = 67$ ) alluded to a beneficial effect of the drug (prevention of neuropathic pain).

### 3.2 HPV Vaccine and Infertility

A total of 2236 posts related to HPV vaccine and infertility were retrieved, again with the majority of posts (85 %) representing data from Twitter (see Table 3). There were 23 posts on Facebook, while 308 posts were retrieved on Google+. Reposts or retweets comprised 23 % of all posts. Similar to posts related to the previous case study on rosiglitazone, the country of origin was unknown for more than half of the HPV vaccine-related posts, with the US representing the majority ( $n = 567$ ) of those posts that could be automatically identified. However, in contrast to the rosiglitazone-related posts, a large proportion of all posts (84 %) referenced other web pages (see Table 4). Various blogs comprised almost half of the linked web pages referenced ( $n = 872$ ), followed by news items ( $n = 669$ ) and scientific journals ( $n = 118$ ). Most of the

**Fig. 1** Trend of assertions of rosiglitazone/cardiovascular event-related posts over time. *EMA* European Medicines Agency, *EU* European Union, *FDA* Food and Drug Administration



**Table 3** Overview of posts about HPV vaccine and infertility across social media networking platforms

Social media platform	No. of posts (%)	No. of reposts (%)	No. of posts with links to other sites (%)	Earliest/latest date of retrieved post	Origin of post (country, based on account holder) <sup>a</sup>
Facebook	23 (1)	6 (26)	15 (65)	April 2014/September 2014	Unknown (19) Bangladesh, India, The Philippines, US (1 each)
Google+	308 (14)	67 (22)	286 (93)	September 2011/September 2014	Unknown (197) US (61) Canada (7) Australia, India (6 each) UK (4) Other countries (27)
Twitter	1905 (85)	437 (23)	1570 (82)	July 2008/September 2014	Unknown (1059) US (505) Canada (112) Australia (40) UK (38) Italy, Egypt (10 each) Other countries (131)
Total	2236	510 (23)	1871 (84)		

<sup>a</sup> Where applicable, only the top five countries are given

blogs commented on these same news items or journal articles. There were 112 posts referring to health information websites intended for health professionals and 49 posts linking to (mostly antivaccine) YouTube videos, while only a minority of posts were associated with lawyer's websites or advertisements ( $n = 24$ ).

The posts demonstrated predominantly affirmative assertion of the association between HPV vaccine and infertility (79 %;  $n = 1758$ ), with posts that negate the association accounting for 4 % ( $n = 85$ ) and neutral posts accounting for the rest. Most neutral posts were asking for further information (particularly with use of the vaccine



during pregnancy), were related to cervical cancer awareness, or were negative comments about the HPV vaccine in general but not directly related to infertility. Figure 2 shows the trend of assertions over time in relation to events in the timeline of the association of interest. The highest peak of affirmative posts occurred in November 2013 when two sisters, aged 20 and 19 years, alleged at a US federal court that Gardasil (trade name of the HPV vaccine) caused them to go into early menopause and become infertile. The build-up to this peak appears to have been triggered by a study describing three young women who presented with secondary amenorrhea following HPV vaccination [17]; this study was first published online at the end of July 2013 (corresponding to the earlier, but smaller, peak in Fig. 2). Many of the posts within the period from August to

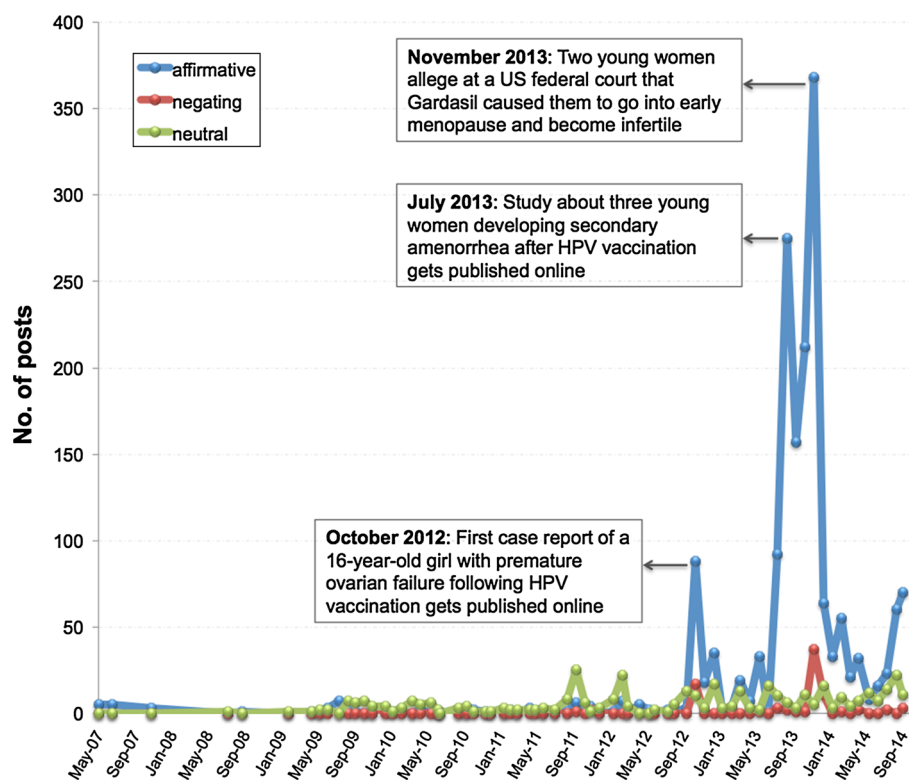
October 2013 actually referred to an event that happened 1 year before—the publication of the first case report on the association of interest. This case report of a 16-year-old Australian girl who had premature ovarian failure after HPV vaccination was first published online in October 2012 [18].

There were nine posts that appeared to be accounts of HPV vaccine–adverse event experience. Six posts involved the person posting the message herself. One simply said she was ‘15 and infertile’ because of the vaccine (the actual page appears to have been taken down after the initial data collection), while four other individuals claimed to have an ovarian cyst, delayed period (and negative pregnancy test), (vaginal) spotting, menopause and hot flashes because of the vaccine. One post was about somebody’s friend who was ‘21 and infertile due to the HPV vaccine’ and there

**Table 4** Description of web pages referenced by posts about human papilloma virus vaccine and infertility

Category of linked web pages	Facebook [n = 15] (%)	Google+ [n = 286] (%)	Twitter [n = 1570] (%)	Total [n = 1871] (%)
News	4 (27)	126 (44)	539 (34)	669 (36)
Law firm’s website or advertisement	–	3 (1)	21 (1)	24 (1)
Blog	5 (33)	111 (39)	756 (48)	872 (47)
Health reference for professionals	–	8 (3)	104 (7)	112 (6)
Scientific journal	–	1 (<1)	117 (7)	118 (6)
Video	1 (7)	16 (6)	32 (2)	49 (3)
Multiple sites, including health education	5 (33)	21 (7)	1 (<1)	27 (1)

**Fig. 2** Trend of assertions of HPV vaccine/infertility-related posts over time. *HPV* human papilloma virus



were two posts from different mothers whose daughters had no (menstrual) periods after receiving the vaccine.

## 4 Discussion

In this study, we aimed to characterise the data currently available from social media networking platforms and to determine if, and how, such data can be tapped for surveillance of two specific safety issues: rosiglitazone and cardiovascular events (i.e. stroke and myocardial infarction), and HPV vaccine and infertility. Rosiglitazone is a drug indicated for a very prevalent disease (diabetes), and although such a disease is expected to occur in the middle-aged population (who comprise a relative minority of the population of Twitter users), it was precisely one of the aims of this study to illustrate that such a group and such a condition of interest could be underrepresented in social media networks, however huge these networks may be. The primary motivation for exploring social media as an additional resource for pharmacovigilance is to capture information that cannot be found in traditional sources. Among the three websites evaluated, Twitter provided the greatest number of (publicly available) posts potentially relevant to the two case studies, but these mostly represented links to news items or, particularly for rosiglitazone and cardiovascular events, websites of personal injury lawyers rather than accounts of drug/vaccine-related adverse events. The ubiquity and instantaneous nature of the Internet and social media networks supposedly provides a mechanism to find adverse drug (or vaccine, or medical device) experiences of laymen that are otherwise missed by ADR reporting systems, and in real time. Thus, one of the more relevant questions to ask is whether data from social media networks can provide early signs of potential safety concerns. Despite the hype about social media representing 'big data', the volume of relevant posts was sparse for the two case studies considered. Although Twitter has over 500 million users (more than half of whom are reportedly active), it was too 'young' a source to use, particularly for the case study on rosiglitazone. When the FDA issued the safety alert on Avandia in May 2007, Twitter had only been in service for less than 1 year, was largely in its trial phase, and thus still had few subscribers. The same argument can be said for Facebook, which became available in September 2006, and Google+, which was launched much later in September 2011. The problem that these social media sites did not have enough time to accumulate data should have been less of an issue for the HPV vaccine–infertility association, which is a more recent potential safety concern, yet that does not seem to be the case.

Our findings corroborate what other researchers have shown regarding the geographic distribution of users of

social media networks: a small number of countries, led by the US, account for a large share of the total user population and likewise make up the active and influential user population [19, 20] (see also <http://www.beevolve.com/twitter-statistics/>). Although this is not totally unexpected, given that only English-language posts were obtained in this study, there can be implications on inferences drawn from research using data from social media networks.

There were (only) 10 and 9 accounts of adverse experiences related to rosiglitazone/cardiovascular events and HPV vaccine/infertility, respectively, but these experiences appeared to be more reactionary than anticipatory (meaning they were shared online after news about the safety issues broke out). Furthermore, verification of such allegations proved to be difficult considering the data privacy constraints (only publicly accessible data could be analysed) and, in particular, establishing an identifiable patient and 'reporter' (required for valid safety reporting in traditional pharmacovigilance systems) is challenging, if not impossible. The scenario of unprincipled individuals spreading inaccurate, and even false, information is not unheard of [21], and since social media is largely unregulated, cannot be avoided. Interestingly, two posts identified in the current study referenced a health information and community website that claims to have studied (as of the time of writing this article) "65,460 people who have side effects while taking Avandia from FDA and social media", and among them, 21,015 had a 'heart attack' (<http://www.ehealthme.com/ds/avandia/heart+attack>). In addition, there were 7752 people who had a 'stroke' (<http://www.ehealthme.com/ds/avandia/stroke>). The website provides statistics on when the heart attack/stroke was reported, age and sex of people who had a heart attack/stroke when taking Avandia, 'time on Avandia when people have a heart attack/stroke', 'severity of the heart attack/stroke when taking Avandia', 'top conditions involved for these people', and 'top co-used drugs for these people'. All such information, if truthful, are relevant. However, nowhere is it stated which part of the information comes from social media and specifically from which social media (there are too many of them). More importantly, there is no description of how these reports were obtained, the actual configuration and content of the reports could not be traced, and the circumstances surrounding the alleged adverse events could not be verified. While the site does include a general disclaimer and a counsel to 'report adverse side effects to the FDA', these sections are found at the end of the page and may be easily ignored.

White et al. [22] utilised retrospective web search logs to make a case for Internet users providing early clues about adverse drug events via their online information seeking. Chary et al. proposed tools for using data from social networks to characterise patterns of (recreational)

drug abuse [23], while Harpaz et al. provided an extensive review on how state-of-the-art text mining for adverse drug events can leverage unstructured data sources, including social media [24]. Similar to the current study, Freifeld et al. used publicly available data from Twitter to obtain messages that resembled adverse event reports ('proto-AEs') related to 23 prespecified medical products [5]. Rather than focusing on a few specific events of interest, the Freifeld et al. study collected all potential events (symptoms), thus resulting in more permutations of search terms, which explains why their study had a higher yield of relevant posts compared with our study. While our current study was more of a 'scoping' study across three social media networking platforms for two specific case studies, the study by Freifeld et al. had a different aim—to evaluate concordance between Twitter posts mentioning AE-like reactions and spontaneous reports received by the FDA Adverse Event Reporting System. There is the implicit assumption of an equivalent level of information between the two sources, which, among other things, necessitated the development of a dictionary to map Internet vernacular to the standardised ontology, Medical Dictionary for Regulatory Activities (MedDRA®). Other researchers have explored the utility of more specific health-oriented websites and patient community forums to identify adverse drug events [25] and to better understand the impact of ADRs [26]. These types of social media sources are likely to provide more relevant content because their very nature allows for sharing of health-related concerns among patients with similar conditions ('like me') and would make verification easier since user registration is often mandatory and more exhaustive (the likelihood of faking an illness in this group is probably lower). Personal accounts of adverse events from such sources are often inaccessible to the public, although many of the prominent and moderated patient community websites will allow access to further information under certain conditions of use (and sometimes for a fee). These more health-oriented social media platforms are certainly worth exploring, especially for surveillance of uncommon adverse events, as well as those related to drugs indicated for rare conditions.

The potential value of mining data from social networks appears to be greatest for measuring awareness regarding potential safety concerns. Because this study focused only on English-language posts, there is the caveat that the findings are biased towards users from English-speaking countries, particularly the US, which comprise the majority of subscribers of these social networking sites. Both number of posts and assertion trend in the two case studies were predominantly driven by events that occurred in the US. Another caveat is that bad news is often more popular than good news. The case report of the 16-year-old girl from Australia who had premature ovarian failure after

HPV vaccination fired up huge comments online, while four studies (published earlier or around the same time) [27–30] that showed no evidence of increased risk for new adverse events, including those related to fertility, were practically ignored.

The other, perhaps even more relevant, question to ask is whether data from social media networks can be used to help corroborate, or refute, potential safety concerns by providing information where there is none. It is time to turn the impressionability of social media as an advantage and leverage it towards bringing balanced and evidence-based information to the Internet and its multitude of users.

Our study has several limitations. Data were queried for co-occurrence of the drug/vaccine of interest and the event of interest within the same post or tweet, which may have limited the number of relevant posts obtained. Similarly, the use of publicly available data and English-language-only posts may have contributed to sampling bias. The assertion analysis conducted may not always reflect the true opinion of the user, the very nature of social media promoting an open and unrestricted environment. A generalisation cannot be made as to which among the social networking platforms provides the most valuable information since the amount and nature of commentaries generated and shared within each network is a function of its own culture and privacy restrictions. Moreover, the population of users of social networking sites comprises the relatively young (and healthy) and fairly educated who have access to the Internet [31–33]. The evaluation undertaken was retrospective and the findings for these particular case studies considered may not necessarily reflect discussions about safety concerns related to other drugs or other vaccines in the future. Because social media platforms are continually being re-engineered to improve the commercial service, there is the concern as to whether studies conducted on data collected from these platforms are reproducible, even 1 year later [34]. The phenomenon of 'blue team dynamics' has been described where the algorithm generating the data (and, consequently, user utilisation) has been modified by service providers such as Google, Twitter and Facebook in line with their business model [34, 35]. Similarly, there are the so-called 'red team' dynamics, which occur when social media platform users attempt to manipulate the data-generating process to support their own economic or political gain [34, 36].

## 5 Conclusions

Publicly available data from the considered social media networks were sparse and largely untrackable for the purpose of providing early clues of safety concerns regarding the prespecified case studies (rosiglitazone and stroke/



myocardial infarction, and HPV vaccine and infertility). The potential value of mining data from social networks appears to be greater for measuring awareness regarding emerging safety issues, with the caveat that this will be biased towards a younger and healthier population who comprise the majority of subscribers of these social networking sites. Further research investigating other case studies (including prospective investigations) and exploring other social media platforms are necessary to further characterise the usefulness of social media for postmarketing safety surveillance.

### Compliance with Ethical Standards

Only publicly available data were used in this study. No ethical approval or informed consent was necessary.

**Funding** No sources of funding were used to assist in the preparation of this study.

**Conflict of interest** Preciosa M. Coloma, Benedikt Becker, Miriam C. J. M. Sturkenboom, Erik M. van Mulligen and Jan A. Kors have no conflicts of interest that are directly relevant to the content of this study.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Adams SA. Sourcing the crowd for health services improvement: the reflexive patient and “share-your-experience” websites. *Soc Sci Med*. 2011;72:1069–76.
- McKee M, Cole K, Hurst L, Aldridge RW, Horton R. The other Twitter revolution: how social media are helping to monitor the NHS reforms. *BMJ*. 2011;342:d948.
- Greene JA, Choudhry NK, Kilabuk E, Shrank WH. Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook. *J Gen Intern Med*. 2011;26:287–92.
- Knezevic MZ, Bivolarevic IC, Peric TS, Jankovic SM. Using Facebook to increase spontaneous reporting of adverse drug reactions. *Drug Saf*. 2011;34:351–2.
- Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, et al. Digital drug safety surveillance: monitoring pharmaceutical products in twitter. *Drug Saf*. 2014;37:343–50.
- Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, et al. Utilizing social media data for pharmacovigilance: a review. *J Biomed Inform*. 2015;54:202–12.
- Goldman SA. Limitations and strengths of spontaneous reports data. *Clin Ther*. 1998;20(Suppl C):C40–4.
- Gonzalez-Gonzalez C, Lopez-Gonzalez E, Herdeiro MT, Figueiras A. Strategies to improve adverse drug reaction reporting: a critical and systematic review. *Drug Saf*. 2013;36:317–28.
- US FDA. For industry: using social media. <http://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/ucm397791.htm>. Accessed 30 Apr 2015.
- US FDA. Guidance for Industry. Internet/social media platforms with character space limitations—presenting risk and benefit information for prescription drugs and medical devices. 2014. <http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm401087.pdf>. Accessed 30 Apr 2015.
- European Medicines Agency. Guideline on good pharmacovigilance practices (GVP). Module VI—Management and reporting of adverse reactions to medicinal products. [http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2014/09/WC500172402.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/09/WC500172402.pdf). Accessed 30 Apr 2015.
- The United States Senate Committee on Finance. Grassley, Baucus release committee report on avandia. 20 Feb 2010. <http://www.finance.senate.gov/newsroom/chairman/release/?id=bc56b552-efc5-4706-968d-f7032d5cd2e4>. Accessed 10 June 2015.
- Graham DJ, Ouellet-Hellstrom R, MaCurdy TE, Ali F, Sholley C, et al. Risk of acute myocardial infarction, stroke, heart failure, and death in elderly Medicare patients treated with rosiglitazone or pioglitazone. *JAMA*. 2010;304:411–8.
- Nissen SE, Wolski K. Rosiglitazone revisited: an updated meta-analysis of risk for myocardial infarction and cardiovascular mortality. *Arch Intern Med*. 2010;170:1191–201.
- Mahaffey KW, Hafley G, Dickerson S, Burns S, Tourt-Uhlig S, et al. Results of a reevaluation of cardiovascular outcomes in the RECORD trial. *Am Heart J*. 2013;166(240–249):e1.
- Takahashi Y, Hasegawa-Moriyama M, Sakurai T, Inada E. The macrophage-mediated effects of the peroxisome proliferator-activated receptor-gamma agonist rosiglitazone attenuate tactile allodynia in the early phase of neuropathic pain development. *Anesth Analg*. 2011;113:398–404.
- Colafrancesco S, Perricone C, Tomljenovic L, Shoenfeld Y. Human papilloma virus vaccine and primary ovarian failure: another facet of the autoimmune/inflammatory syndrome induced by adjuvants. *Am J Reprod Immunol*. 2013;70:309–16.
- Little DT, Ward HR. Premature ovarian failure 3 years after menarche in a 16-year-old girl following human papillomavirus vaccination. *BMJ Case Rep*. 2012;2012. doi:10.1136/bcr-2012-006879.
- Kulshrestha J, Kooti F, Nikraves A, Gummadi KP. Geographic dissection of the Twitter network. Association for the Advancement of Artificial Intelligence. 2012. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4685>. Accessed 10 June 2015.
- Sullivan SJ, Schneiders AG, Cheang CW, Kitto E, Lee H, et al. ‘What’s happening?’ A content analysis of concussion-related traffic on Twitter. *Br J Sports Med*. 2012;46:258–63.
- Edwards IR, Lindquist M. Social media and networks in pharmacovigilance: boon or bane? *Drug Saf*. 2011;34:267–71.
- White RW, Tatonetti NP, Shah NH, Altman RB, Horvitz E. Web-scale pharmacovigilance: listening to signals from the crowd. *J Am Med Inform Assoc*. 2013;20:404–8.
- Chary M, Genes N, McKenzie A, Manini AF. Leveraging social networks for toxicovigilance. *J Med Toxicol*. 2013;9:184–91.
- Harpaz R, Callahan A, Tamang S, Low Y, Odgers D, et al. Text mining for adverse drug events: the promise, challenges, and state of the art. *Drug Saf*. 2014;37:777–90.
- Chee BW, Berlin R, Schatz B. Predicting adverse drug events from personal health messages. *AMIA Annu Symp Proc*. 2011;2011:217–26.
- Butt TF, Cox AR, Oyeboe JR, Ferner RE. Internet accounts of serious adverse drug reactions: a study of experiences of Stevens-Johnson syndrome and toxic epidermal necrolysis. *Drug Saf*. 2012;35:1159–70.

27. Slade BA, Leidel L, Vellozzi C, Woo EJ, Hua W, et al. Postlicensure safety surveillance for quadrivalent human papillomavirus recombinant vaccine. *JAMA*. 2009;302:750–7.
28. Agorastos T, Chatzigeorgiou K, Brotherton JM, Garland SM. Safety of human papillomavirus (HPV) vaccines: a review of the international experience so far. *Vaccine*. 2009;27:7270–81.
29. Gee J, Naleway A, Shui I, Baggs J, Yin R, et al. Monitoring the safety of quadrivalent human papillomavirus vaccine: findings from the Vaccine Safety Datalink. *Vaccine*. 2011;29:8279–84.
30. Klein NP, Hansen J, Chao C, Velicer C, Emery M, et al. Safety of quadrivalent human papillomavirus vaccine administered routinely to females. *Arch Pediatr Adolesc Med*. 2012;166:1140–8.
31. Chou WY, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. *J Med Internet Res*. 2009;11:e48.
32. Nef T, Ganea RL, Muri RM, Mosimann UP. Social networking sites and older users: a systematic review. *Int Psychogeriatr*. 2013;25:1041–53.
33. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. *J Med Internet Res*. 2015;17:e70.
34. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. *Science*. 2014;343:1203–5.
35. Butler D. When Google got flu wrong. *Nature*. 2013;494:155–6.
36. Mustafaraj E, Metaxas P. From obscurity to prominence in minutes: political speech and real-time search. *Proceedings of the WebSci10: Extending the Frontiers of Society On-Line*, Raleigh, NC, 26–27 April 2010.